REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*

2. REPORT DATE
July 1, 1998

3. REPORT TYPE AND DATES COVERED
Final Progress Report, 5/95-4/98

4. TITLE AND SUBTITLE

Multimedia, Multipoint Communications

5. FUNDING NUMBERS

DAAH04-95-1-0188

6. AUTHOR(S)

Professors Mischa Schwartz, Dimitris Anastassiou, Alexandros Eleftheriadis

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

Columbia University
Office of Projects and Grants
351 Engineering Terrace
New York, NY 10027

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 33077.6-EL

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12.b DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This research project focused on two parallel, complementary research activities designed to address potential problems arising in the transmission of multimedia traffic to multiple parties over both high-speed wired and lower speed wireless communication networks of the future. The first addressed appropriate flow/congestion control in this multicast environment, and specifically the design of algorithms to control the flow of traffic between a source and multiple receivers, either fixed or mobile, located at possibly different distances from the source, so that the receivers can each handle the traffic flow, and yet provide as high a throughput as possible with queueing delays constrained to acceptable values. The second activity addressed the efficient transmission of digital, standards-based video (MPEG-1/MPEG-2) in this networking environment. To mitigate the highly variable nature of network conditions in terms of available bandwidth and delay, we investigated algorithms that combined rate shaping (on-the-fly modification of the video bit rate) and flow control to maximize the received video quality. To ensure that the video traffic is a fair and equitable user of the network, we used TCP flow control but no error control, so that video competes fairly for resources with traditional types of traffic (which are predominantly TCP-based).

14. SUBJECT TERMS
Multipoint communications, multipoint flow control, rate shaping, Internet video, streaming

15. NUMBER OF PAGES
7

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

DTIC QUALITY INSPECTED 5

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

Multimedia, Multipoint Communications

Final Progress Report

Prepared by
Profs. Mischa Schwartz and Dimitris Anastassiou
Principal Investigators

and
Prof. Alexandros Eleftheriadis

July 1, 1998

U.S. Army Research Office

Grant No. DAAH04-95-1-0188

Columbia University in the City of New York

Approved for Public Release; Distribution Unlimited

1. Statement of Problem Studied

This research project focused on two parallel, complementary research activities designed to address potential problems arising in the transmission of multimedia traffic to multiple parties over both high-speed wired and lower speed wireless communication networks of the future.

The first area was that of providing appropriate flow/congestion control in this multicast environment. Specifically, we have addressed the objective of designing good algorithms to control the flow of traffic between a source and multiple receivers, either fixed or mobile, located at possibly different distances from the source, so that the receivers can each handle the traffic flow, and yet provide as high a throughput as possible, with queueing delays constrained to acceptable values.

The second area of research involved the efficient transmission of digital, standards-based video in this networking environment. Recognizing the highly variable nature of network conditions in terms of available bandwidth and delay, we investigated novel algorithms that would combine rate shaping (on-the-fly modification of the bit rate of MPEG-1 and MPEG-2 video) and flow control strategies to maximize the received video quality. In doing so, we focused on techniques that would make the video traffic a fair and equitable user of the network, in the sense that it would compete fairly for resources with other traditional types of traffic (more specifically, the TCP-based SMTP, HTTP and FTP traffic which account for most of the traffic on the Internet today).

2. Summary of the Most Important Results

We summarize here the results in the two research areas studied.

2.1 Multicast Flow and Congestion Control Over Combined Wired/Wireless Networks

Multipoint multimedia communication is becoming more and more important in today's Internet environment as well as in more futuristic mixed wired/wireless network environments. The delivery of multicast information over networks which are designed primarily for point-to-point communications poses interesting management and control issues that need to be resolved. Flow and congestion control is one such challenging problem.

In this portion of our work we proposed and studied a unified framework for the design and performance analysis of multicast flow control algorithms, as well as some possible source policies, such as listening to the slowest request, random listening, weighted sum, estimation/prediction, etc. Given the complexity of the problem, we first addressed the simplest model with two mobile receivers, using as the source control simple binary on-off rate control.

The receivers were each assumed to be connected by a wireless link to a base station. The wireless link itself was modeled as a two-state Markov chain, switching randomly from on to off, to account for random fading and randomly-occurring severe error conditions. The base station, in turn, was modeled as a queue driving the on-off link model. Simple messages were assumed to be transmitted from the base station back to the source, depending on the algorithm proposed, indicating the state of the queue, whether above or below a specified threshold; or the actual queue size itself, as well as rate of change of the queue. The base stations were each located at different distances, hence incurring different propagation delays, away from the source. The receiver characteristics were chosen to be similar (the homogeneous case) or different (the heterogeneous case).

Three ad hoc algorithms were proposed and studied comparatively. These included, first, the Listen to the Slowest Request (LSQ) algorithm, in which each receiver sends a single-bit "start-stop" feedback signal: a "stop" signal whenever the queue length exceeds a specified high threshold; a "start" signal whenever the queue length exceeds a specified low threshold. The source stops sending traffic whenever at least one

receiver has sent a "stop" signal. It sends at a constant rate otherwise. The second, more complex, algorithm was the Source Estimation (SE) algorithm, in which each receiver sends back both its queue length and the queue growth rate, whenever that rate changes sign. The source then estimates the future queue length of each receiver, taking the propagation delay into account. It transmits at a constant rate only if at least one estimated queue length drops below a specified low threshold and all the other queues stay below specified high thresholds. Otherwise the source stops transmitting. The third control algorithm used for comparison was a simple open loop control, in which the source collects from each receiver at call setup time information as to its location, its desired average rate of reception of data packets, its peak rate, and desired quality of service requirements, among other parameters. The source then adjusts its rate of transmission to satisfy those parameters, with no signals fed back.

The three algorthms were compared, in the general case, using simulation. We were able to obtain analytic results in the case of zero propagation delay. These results were found to agree qualitatively with the simulation results with propagation delay included. The results indicated that the more complex SE algorithm was able to provide better delay-throughput performance than the LSQ algorithm. However, as the propagation delays increased the relative performance improvement was reduced, since prediction of queue lengths was increasingly less accurate. For rapid link fading rates the open-loop algorithm becomes a good policy to use, with the feedback signals providing relatively little added information. Various aspects of this work have been presented at conferences and published in technical journals. Details appear in references [1], [2], [3], [4] in the list of publications that follows.

After studying the two-receiver system with binary on-off control, we went on to study more complex rate-based flow control algorithms. The ATM Forum has selected a rate-based flow control mechanism to support its projected ABR service in the point-to-point, single receiver case, so it is of considerable interest to study multicast rate-based flow control in a more general case than the on-off control case. We focused in this work on the fixed receiver, high-speed wired environment only,

Specifically, we studied a "weighted sum" (WS) algorithm, which appears suited to the multicast environment, and analytically proved it to be stable under proper parameter settings. This algorithm extends earlier work on the stability of a unicast feedback flow control algorithm carried out at Bell Labs. In this weighted sum extension of the unicast algorithm, the source picks an appropriate gain factor with which to weight the (propagation-delayed) queue size information received from each receiver involved in the multicast tree and sums the weighted information in determining the appropriate transmission rate to use. An additional damping coefficient is added to ensure stability. A detailed analysis was carried out of the two-receiver case, using a fluid-flow approach to model the resultant system. The model results in a second-order deterministic delay-differential equation to be solved. In general, this is a very difficult problem to attack. We were able, however, to use a stability approach technique developed by a researcher at NEC C&C Labs, Princeton, to carry out the stability analysis in this case, and come up with rules for determining the optimal parameter settings to not only ensure a stable systgem, but to provide the fastest, oscillation-free transient behavior. Details of this work appear in [1] and [5].

The third and final portion of this work on multicast flow control involved window-based flow control in the Internet environment. The objective was to come up with a TCP-like multicast flow control algorithm that retained as many of the TCP unicast flow control features as possible, including being responsive to congestion, yet was fair in allocating bandwidth to all users using the same network links. The fairness issue had two components: the multicast traffic had to be fair in sharing bandwidth with TCP along the same paths; multiple multicast sessions between the same sender and receivers had to share the bandwidth equally, on average.

To accomplish these goals, we came up with a congestion-control algorithm we call the "Random Listening Algorithm", which cuts the window at the source on a random basis when receiving notification of congestion from any receiver. The rationale here was two-fold: first, occasional packet losses do not necessarily mean congestion is present; second, given a congested ("bottleneck") link along a source-

receiver path, multiple congestion messages might be received from receivers affected. One would like the source to respond only once in this case, not to each such message received. The algorithm was thus designed to only respond to receivers reporting frequent losses. On receiving a congestion message from such a receiver, the sender reduces its window with a probability 1/n, where n is the number of such receivers. Details of the algorithm appear in [1] and [6].

Both analysis and simulation were used to show the algorithm was stable and fair. To prove the latter we had to define a quantitative measure of fairness we call "essentially fair", which provides a lower and upper bound on the multicast throughput with respect to any TCP connections using the same source-receiver paths. We then proved fairness of the algorithm along a restricted tree-based topology for various congestion cases, for both drop-tail and RED gateways along the path. (Drop-tail gateways are extensively used in the current Internet; RED gateways have been proposed to reduce network queueing delay and to provide a fairer response to congestion.) Extensive simulations were carried out to both verify our analysis and to validate the effectiveness of the random listening algorithm.

In addition to the work described above on multicast flow control for wired and combined wired/wireless networks, work was carried out on multimedia admission control in cellular mobile wireless networks. A distributed predictive admission control algorithm was developed that provides appropriate throughput to the system while maintaining a fixed call dropping probability for each traffic class, as defined by the user. The algorithm is based on simple Markov analysis and traffic prediction. It was shown, by simulation, to compare favorably with other admission control algortihms proposed in the literature. This work was presented at an international Workshop on digital communications and has been published, in book form, as part of the Workshop Proceedings [7].

2.2 Digital Video and Audio Processing for Multipoint, Multimedia Communications

In this portion of our work we focused on techniques that would adapt the rate of compressed digital video and audio information in order to accommodate the dynamic nature of available bandwidth and delay in both single and multi-point communications. Our approach consisted of combining signal processing principles and flow control strategies in order to maximize the perceived quality of the information received by the users while operating within realistic network performance parameters. In addition, we focused on well-known audio and video coding techniques, more specifically MPEG-1 and MPEG-2.

The problem was decomposed into two parts. First, a method had to be developed to allow the modification of the bit rate of compressed digital audio and video to arbitrary and dynamically varying constraints. Several techniques have been reported in the literature in the context of traditional rate control, including some that take into account network conditions (e.g., by Reibman et al. in the former AT&T Bell Labs). These, however, require that the encoder is available and is tightly coupled with the network. This precludes application of these techniques on precompressed, stored content. Further, it requires substantial modification on the encoding algorithm, something not easily done with high-quality hardware-based encoders.

To mitigate these problems, we used a refined version of our dynamic rate shaping technique, which provides for arbitrary bit rate modification of compressed video according to user-imposed constraints. Rate shaping works on any block-based transform codec (including MPEG-1/MPEG-2, H.261, H.263, M-JPEG). Its principle lies in the optimal removal of transform (typically, Discrete Cosine Transform – DCT) coefficients from a compressed bitstream prior to its transmission on the network. Optimallity here is in the context of minimizing the unavoidable distortion for the desired target bit rate. We have developed the mathematical framework for the analysis of this problem, and derived optimal algorithms as well as fast approximations that are within 0.5 dB of the optimal one. A particularly attractive feature of the fast approximations are that they are still near-optimal even if error propagation is ignored. This

propagation is generated by the recursive nature of video coding (i.e., coding of P and B frames, in MPEG parlance). The overall complexity of these algorithms is much less than a fulll decoder, thus making them suitable for software-only implementations. The achieved bit rate reduction depends on the original resolution and bit rate of the video signal, and can be as much as 50% or more.

Our prior work in this area only considered the mathematical problem, and only fixed target bit rates. In order to apply this technique to a dynamic network environment, it was essential to extend it to handle time-varying constraints. A particularly challenging problem is the origin of these constraints: in order for them to be useful, they must faithfully represent the actual available bandwidth on the given network. Our efforts concentrated on the Internet, as a very relevant example of a dynamically-varying network. While there are several techniques for estimating network bandwidth, an additional desirable characteristic that we identified was fairness. In other words, we would want to ensure that the video traffic would behave inside the network in exactly the same way as traditional data traffic, but without ignoring video's special characteristics (tolerance for losses, need for low delay and jitter). Considering that the vast majority of the Internet traffic today is based on TCP (which is the transport protocol used in HTTP, FTP, and SMTP), we decided to use TCP's window-based flow control but without error control. The benefit of this approach is that the injected traffic behaves as any other TCP-based traffic, while the delays incurred by the retransmissions imposed by error control are eliminated. By definition, this appoach guarantees fairness, something that is very hard to ascertain for any other scheme due to the non-linearity and complexity of the TCP algorithms.

This window-based flow control, however, cannot be directly applied to the rate shaper. The latter requires an actual bit budget in order to modify the number of bits to be used on each frame. We devised a solution that uses the traditional buffer-constrained rate control model used in video and audio coding systms. A buffer is placed between the rate shaper and the TCP flow control algorithm; the rate shaper modifies the target bit rate (and thus, bit budget per frame) depending on the occupancy of the buffer. The buffer, in turn, is emptied at the rate specified by the flow control algorithm. The benefit of this technique is that it can provide for very stable control of the bit rate while avoiding rapid fluctuations and the corresponding undesirable effects on video quality.

A complete system was implemented in software using a Sun workstation as a server and Windows 95 as the client. The server contained an optimized implementation of a rate shaper as well as an implementation of the TCP flow control algorithm in user-mode (i.e., outside the OS kernel) operating over UDP. On the client side, software-based video decoding was used based on Microsoft's ActiveMovie environment (renamed to DirectShow). In particular, an ActiveMovie filter was developed to implement the client-side of the flow control algorithm. The system is capable of operating in real-time, providing video rates up to 30 frames per second for frame sizes of 160x120. Extensive tests were performed both on our local area network, and also across the actual Internet in connections extending over 30 hops. As a result of these tests, some additional features were added, such as selective rentrasmission when the end-to-end delay allows it (to increase robustness). Subjective comparison with several commercial solutions (e.g., by VDOnet) demonstrated that our technique outperformed them (better quality, higher frame rate). A particularly important feature is that the rate adaptation allows much higher frame rates to be used, thus substantially improving the user's experience. Results of this work have been reported in numerous journal publications, conference presentations, and demonstrations (see [9], [10], [13]-[17]. In addition, a US patent has been filed (see Section 5 below).

We also examined the use of such a system over wireless networks. The fundamental difference lies in TCP's behavior over wireless links. TCP performance over wireless networks can be poor because TCP assumes losses occur in the network only when there is congestion. However, wireless networks often drop packets due to random errors on the wireless link. TCP performance can be severely degraded when packets with errors are mistaken for congestion. We devised a very simple, tractable, and closed-form analytical model for determining the performance of TCP over hybrid wired/wireless networks. We compared our model with previous work in this area from an analytical perspective. We also compared

our model with several simulations. The results showed that our model matches both the previous work and the simulations very well. We also provided a simple heuristic for determining when wireless errors will severely degrade TCP performance. These results are reported in [11].

In the course of this investigation, we also had to address video server scalability issues. Since rate shaping requires computational resources from the server, it has a non-trivial impact on the number of streams it can simultaneously serve. We investigated the performance aspects of high-performance video pumps in our Video-on-Demand testbed, and identified the PDU (protocol data unit) size as one of the most significant problems in server scalability. In particular, we demonstrated that a smaller PDU size (mandated by buffering resctrictions on simple receivers such as set-top boxes) has a devastating impact on pumps based on general-purpose workstations (even in high-end systems such as the Silicon Graphics Onyx/Challenge). Our results are reported in [12] and [17].

Our experiments with the above techniques have concentrated almost exclusively on video. However, the same principles apply equally well on audio. For the case of speech signals, there are additional considerations that simplify its treatment in a multimedia networking environment. In particular, the fact that silence detection can be used prior to transmission to eliminate silent signal parts can be of great use to mitigate delay and bandwidth variations on a network. This is similar in spirit to rate shaping, but the rate adaption utilizes perceptual criteria rather than being applied to the entire signal. We developed a novel silence detection scheme for the explicit use in a general-purpose computing environment. By integrating in our Xphone videoconferencing system (see A. Eleftheriadis et al., "Architecture and Algorithms of the Xphone Multimedia Communication System," *ACM Multimedia Systems Journal*, Vol. 2, No. 2, August 1994, pp. 89-100), we demonstrated that it provides an excellent tradeoff between computational simplicity and accuracy. Detailed results and comparisons with various state-of-the-art schemes are presented in [8].

3. List of Publications

[1] Huayan Amy Wang, **Multicast Flow and Congestion Control Over Combined Wired/Wireless Networks**, Ph.D. Dissertation, Columbia University, New York, NY, 1998

[2] H. Wang and M. Schwartz, "Comparison of Multicast Flow Control Algorithms over Combined Wired/Wireless Networks," in **Mobile Multimedia Communications**, D. J. Goodman and D. Raychaudhuri, eds, Plenum Press, NY, 1997, pp. 91-100. (*Proc., 3rd International Workshop on Mobile Multimedia Communications*, MoMuc-3, Sept. 1996, Princeton, NJ).

[3] H. Wang and M. Schwartz, "Performance Analysis of Multicast Flow Control Algorithms over Combined Wired/Wireless Nertworks," *IEEE INFOCOM'97*, Kobe, Japan, April 1997.

[4] H. Wang and M. Schwartz, "Performance Analysis of Multicast Flow Control Algorithms over Combined Wired/Wireless Networks," *IEEE J. on Selected Areas on Communications*, vol. 15, no. 3, Sept. 1997, pp. 1349-1363.

[5] H. Wang, R. Izmailov, and M. Schwartz, "Adaptive Rate-Based Feedback Flow Control Algorithms-Extensions to Multicast Connections," *Proc. ISCC'98*, Greece, June 1998.

[6] H. Wang and M. Schwartz, "Achieving Bounded Fairness for Multicast and TCP Traffic in the Internet," *SIGCOMM'98*, Vancouver, Canada, Sept. 1998.

[7] B.Epstein and M. Schwartz, "QoS-based Predictive Admission Control for Multimedia Traffic," **Broadband Wireless Communications**, M. Luise and S. Pupolin, eds, Springer, London, 1998. (*Proc., International Tyrrhenian Workshop on Digital Communications*, Lerici, Italy, Sept. 1997.)

[8] S. Jacobs, A. Eleftheriadis, and D. Anastassiou, "Silence Detection for Multimedia Communication Systems," *ACM/Springer Verlag Multimedia Systems Journal*, 1998 (to appear).

[9] S. Jacobs and A. Eleftheriadis, "Streaming Video using TCP Flow Control and Dynamic Rate Shaping," *Journal of Visual Communication and Image Representation*, Special Issue on Image Technology for World-Wide-Web Applications, 1998 (to appear).

[10] S. Jacobs and A. Eleftheriadis, "A Real-Time Protocol that Guarantees Fairness with TCP," submitted to *ACM/Springer Verlag Multimedia Systems Journal*, July 1998.

[11] S. Jacobs, Huayan Wang, A. Eleftheriadis, and M. Schwartz, "A Simple Analytical Model for Wireless TCP Utilization," submitted to *ACM/Baltzer Science Wireless Networks Journal*, July 1998.

[12] S. Jacobs and A. Eleftheriadis, "Video Pump Design for Interoperability with Set Top Units: The Case Against Small PDUs," *Proceedings, 6th International Workshop on Network and Operating System Support for Digital Audio and Video* (NOSSDAV-96), Zushi, Japan, April 1996, pp. 123–129.

[13] S. Jacobs and A. Eleftheriadis, "Providing Video Services over Networks without Quality of Service Guarantees," *Proceedings, WWW Consortium Workshop on Real-Time Multimedia and the Web*, Sophia Antipolis, France, October 1996.

[14] S. Jacobs and A. Eleftheriadis, "Adaptive Video Applications for Non-QoS Networks," *Proceedings, IFIP Fifth International Workshop on Quality of Service*, May 1997.

[15] S. Jacobs and A. Eleftheriadis, "Real-Time Dynamic Rate Shaping and Control for Internet Video Applications", *Proceedings, 1st IEEE Workshop on Multimedia Signal Processing*, June 1997 (invited paper and demonstration).

[16] S. Jacobs and A. Eleftheriadis, "Real-Time Video on the Web using Dynamic Rate Shaping," *Proceedings, 4th IEEE International Conference on Image Processing* (ICIP-97), Santa Barbara, California, October 1997.

[17] S. Jacobs, **Media and Protocols for Multimedia Communications over Disparate Networks,** Ph.D. Dissertation, Columbia University, New York, NY, 1998.

4. List of All Participating Scientific Personnel

4.1 Faculty

Prof. Mischa Schwartz
Prof. Dimitris Anastassiou
Prof. Alexandros Eleftheriadis

4.2 Graduate Research Assistants

Huayan Amy Wang, earned Ph.D. degree, 1998
Bracha Epstein
Stephen E. Jacobs, earned Ph.D. degree, 1998

5. Report of Inventions

S. Jacobs and A. Eleftheriadis, "Transmission Control for Minimizing Congestion in Digital Communication Networks," U.S. Patent Application, filed October 24, 1997.